Finding Hedges by Chasing Weasels: Hedge Detection Using Wikipedia Tags and Shallow Linguistic Features

Viola Ganter and Michael Strube
EML Research gGmbH Heidelberg, Germany
ACL and AFNLP, 2009

22 May, 2014
Jaehwan Lee
Outline

- Introduction
- Related Work
- Weasel Words
- Data and Annotation
- Method
  - Words Preceding Weasel Tags
  - Adding shallow linguistic features
- Results and Discussion
- Conclusions
Introduction

- Distinguishing facts from fiction

- Indicate that speakers do not back up their opinions with facts

- in Abstract
  - “We investigate the automatic detection of sentences containing linguistic hedges using corpus statistics and syntactic patterns
Introduction

- Distinguishing facts from fiction

- Indicate that speakers do not back up their opinions with facts

- in Abstract
  - “We investigate the automatic detection of sentences containing linguistic hedges using **corpus statistics** and **syntactic patterns**"
Related Work

- Focused on the biomedical domain
  - Light et al. (2004)

- Weakly supervised system for hedge classification
  - in a very narrow subdomain in the life sciences
  - Medlock and Briscoe (2007)
Weasel Words

- Wikipedia editors are advised to avoid *weasel words*
  - E.g. “Some people say …”, “I think …”, “Clearly …”

- Wikipedia style guidelines instruct editors to
  - if they notice weasel words, insert a {{weasel-inline}} or a {{weasel-word}} tag to mark sentences or phrases for improvement
  - E.g. Others argue {{weasel-inline}} that the news media are simply catering to public demand.

- Many Wikipedia articles contain a specific weasel tag
  - so that Wikipedia can be viewed as a *readily annotated corpus*
Data and Annotation

- Balanced set
  - chose one random, non-tagged sentence per tagged sentence
  - Wikipedia dumps from years 2006 to 2008
    - articles that contained the string {{weasel
    - 168,923 unique sentences containing 437 weasel tags
    - one random, non-tagged sentence per tagged sentence
    - resulting in a set of 500 sentences
  - Wikipedia dumps completed on March 6, 2009
    - 70,436 sentences with 328 weasel tags
    - Again, a balanced set of 500 sentences

- Manually annotated set
  - expected there to be a much higher number of potential weasel words which had not yet been tagged leading to false positives
  - one of the authors, two linguists and one computer scientist
  - resulting in a set of 246 sentences for evaluation
Method

- in Abstract
  - “We investigate the automatic detection of sentences containing linguistic hedges using **corpus statistics and syntactic patterns**

- Corpus statistics
  - Words Preceding Weasel Tags (wpw)

- Syntactic patterns
  - Adding shallow linguistic features (asp)
Method
Words Preceding Weasel Tags (wpw)

- Assumption
  - weasel phrases contain at most five words
  - weasel tags are mostly inserted behind weasel words or phrase

\[ \cdots \{\text{weasel tag}\}\cdots \]
### Assumption
- Weasel phrases contain at most five words.
- Weasel tags are mostly inserted behind weasel words or phrases.

Weasel words are prefixed by colored boxes to indicate their weasel status.
Method

Words Preceding Weasel Tags (wpw)

- Assumption
  - weasel phrases contain at most five words
  - weasel tags are mostly inserted behind weasel words or phrase

- Two Factors
  - Relative frequency
  - Average distance

- Equation goes,

\[
Score(w) = RelF(w) + AvgDist(w) \quad (1)
\]
Method

Words Preceding Weasel Tags (wpw)

- Relative frequency

\[ \text{RelF}(w) = \frac{W(w)}{\log_2(C(w))} \] (2)

- \(W(w)\): the number of times word \(w\) occurred in the context of a weasel tag
- \(C(w)\): the total number of times \(w\) occurred in the corpus
- to give those words a high score which occur frequently in the context of a weasel tag

- due to the sparseness of tagged instances, words that occur with a very high frequency in the corpus automatically receive a lower score than low-frequency words
- Thus, use the logarithmic function to diminish this effect
Method

Words Preceding Weasel Tags (wpw)

- Average distance

\[ \text{AvgDist}(w) = \frac{W(w)}{\sum_{j=0}^{W(w)} \text{dist}(w, \text{weaseltag}_j)} \]  

(3)

- \( j \): each weasel context
- \( \text{dist}(w, \text{weaseltag}_j) \): the distance of word \( w \) to the weasel tag in \( j \)

- E.g. A word that always appears directly before the weasel tag will receive an \( \text{AvgDist} \) value of 1
- E.g. A word that always appears five words before the weasel tag will receive an \( \text{AvgDist} \) value of \( 1/5 \)
Method

Words Preceding Weasel Tags (wpw)

- **Normalization**
  - \( wpw(S) \): the sum of scores over all words in \( S \)
  - normalized by the hyperbolic tangent

\[
wpw(S) = \tanh \sum_{i=0}^{\|S\|} \text{Score}(w_i) \quad (5)
\]

with \( |S| = \) the number of words in the sentence.

- **Classification**
  - After calculating \( wpw(S) \) score for a sentence \( S \)
  - if \( wpw(S) \) is larger than a threshold, it is classified as weasel

\[
S \rightarrow \text{weasel if } wpw(S) > \sigma \quad (4)
\]
Method

Adding Shallow Linguistic Features (asp)

- the Weasel words in Wikipedia can be divided into
  - Numerically underspecified subjects ("Some people", "Many")
  - Passive constructions ("It is believed", "It is considered")
  - Adverbs ("Often", "Probably")

- If a pattern is found,
  - only the head of the pattern is assigned a score
  - i.e. adverbs, main verbs for passive patterns, nouns and quantifiers for numerically underspecified subjects

\[
asp(S) = \tanh \sum_{i=0}^{heads_S} \text{Score}(w_i) \tag{6}
\]

where \(heads_S\) = the number of pattern heads found in sentence \(S\).
Results and Discussion

- **Result**
  - Both model perform comparably well on the development test data
  - the syntactic patterns do not contribute to the regeneration of weasel tags
  - Word frequency and distance to the weasel tag are sufficient

- **Limitation**
  - decreasing precision of both approaches when trained on more tagged sentences (i.e., computed with a higher threshold) might be caused by the great number of unannotated weasel words
  - A disadvantage of the weasel tag is its short life span
Results and Discussion

Comparison

– difference becomes more distinct when manually annotated data form the test set
– *asp* out performs *wpw* by a large margin (also because *wpw* performs rather poorly)
– suggests that the added syntactic patterns indeed manage to detect weasels that have not yet been tagged

<table>
<thead>
<tr>
<th></th>
<th>.60</th>
<th>.70</th>
<th>.76</th>
<th>.80</th>
<th>.90</th>
<th>.98</th>
</tr>
</thead>
<tbody>
<tr>
<td>balanced set</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>wpw</em></td>
<td>.68</td>
<td>.68</td>
<td>.68</td>
<td>.69</td>
<td>.69</td>
<td>.70</td>
</tr>
<tr>
<td><em>asp</em></td>
<td>.67</td>
<td>.68</td>
<td>.68</td>
<td>.68</td>
<td>.61</td>
<td>.59</td>
</tr>
<tr>
<td>manual annot.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>wpw</em></td>
<td>-.</td>
<td>.59</td>
<td>-.</td>
<td>-.</td>
<td>-.</td>
<td>.59</td>
</tr>
<tr>
<td><em>asp</em></td>
<td>.68</td>
<td>.69</td>
<td>.69</td>
<td>.69</td>
<td>.70</td>
<td>.65</td>
</tr>
</tbody>
</table>

Table 2: F-scores at different thresholds (bold at the precision/recall break-even-points determined on the development data)
Conclusions

- Main Idea
  - to use Wikipedia as a readily annotated corpus

- The experiments show that
  - the syntactic patterns work better on manual annotations
  - word frequency and distance work better on Wikipedia weasel tags itself

- This approach
  - takes a much broader domain than previous work
  - easily be applied to different languages by using Wikipedia
  - using the Wikipedia edit history will resolve short span of weasel tags